JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Review of Criminal Behavior Analysis based on Machine Learning Techniques

[1]Shrutika Gaikwad, [2]Prof. Anjul Rai
[1]Research Scholar, [2]Assistant Professor
[1&2]Department of Computer Science Engineering,
[1&2]School of Research & Technology, People's University, Bhopal, India

Abstract : Crimes are increasing with a high frequency rates in this new era of world and hence it's a devastating issue that everyone has been experiencing. For finding a pattern that can be used for prediction is necessary. The extraction of new information is predicted using the existing datasets. Many approaches for analysis and prediction in data mining had been performed. But, many few efforts has made in the criminology field. Many few have taken efforts for comparing the information all these approaches produce. The police stations and other similar criminal justice agencies hold many large databases of information which can be used to predict or analyze the criminal movements and criminal activity involvement in the society. The criminals can also be predicted based on the crime data. The main aim of this work is to perform a survey on the machine learning techniques that has been applied towards criminal identification.

IndexTerms **- Crime, Machine Learning, Data Mining, Criminal.**

I. INTRODUCTION

The Crime investigators heavily rely on their large knowledge of criminal behavior. When investigating a new case, applying this knowledge can lead to cognitive overload and tunnel vision. Some support systems are developed to search through historical data and knowledge more easily, but still require the investigator to put all the pieces together and ask the right question. Criminology is an area that focuses the scientific study of crime and criminal behavior and is a process that aims to identify crime characteristics. It is one of the most important fields where the application of data mining techniques can produce important results. A broad analysis of unlawful activity reveals that all criminal behavior shares a common set of universal principles. A micro simulation model can be drawn out by interlinking the universal principles with the attributes of the individuals for profiling the criminal behavior.

Computer vision is a branch of artificial intelligence that trains the computer to understand and comprehend the visual world, and by doing so, creates a sense of understanding of a machine's surroundings [1, 2]. It mainly analyzes data of the surroundings from a camera, and thus its applications are significant. It can be used for face recognition, number plate recognition, augmented and mixed realities, location determination, and identifying objects [3].

Figure 1: Location Map

Research is currently being conducted on the formation of mathematical techniques to recover and make it possible for computers to comprehend 3D images. Obtaining the 3D visuals of an object helps us with object detection, pedestrian detection, face recognition, Eigen faces active appearance and 3D shape models, personal photo collections, instance recognition, geometric alignment, large databases, location recognition, category recognition, bag of words, part-based models, recognition with segmentation, intelligent photo editing, context and scene understanding, and large image collection and learning, image searches,

recognition databases, and test sets. These are only basic applications, and each category mentioned above can be further explored. VL Feat is studied, which is a library of computer vision algorithms that can be used to conduct fast prototyping in computer vision research, thus enabling a tool to obtain computer vision results much faster than anticipated. Considering face detection/human recognition [5], human posture can also be recognized. Thus, computer vision is extremely attractive for visualizing the world around us.

Machine learning (ML) is an application that provides a system with the ability to learn and improve automatically from past experiences without being explicitly programmed [6]. After viewing the data, an exact pattern or information cannot always be determined. In such cases, ML is applied to interpret the exact pattern and information. ML pushes forward the idea that, by providing a machine with access to the right data, the machine can learn and solve both complex mathematical problems and some specific problem.

In general, ML is categorized into two parts: (1) supervised ML and (2) unsupervised ML. In supervised learning, the machine is trained on the basis of a predefined set of training examples, which facilitates its capability to obtain precise and accurate conclusions when new data are given. In unsupervised learning, the machine is given a set of data, and it must find some common patterns and relationships between the data its own. Neural networks, which are important tools used in supervised, learning, have been studied. Although various crimes and their underlying nature seem to be unpredictable.

Nowadays, criminal intellect with the help of advances in technology is improving with each passing year. Consequently, it has become necessary for us to provide the police department and the government with the means of a new and powerful machine (a set of programs) that can help them in their process of solving crimes. The main aim of crime forecasting is to predict crimes before they occur, and thus, the importance of using crime forecasting methods is extremely clear. Furthermore, the prediction of crimes can sometimes be crucial because it may potentially save the life of a victim, prevent lifelong trauma, and avoid damage to private property. It may even be used to predict possible terrorist crimes and activities.

II. LITERATURE SURVEY

G. Jha et al.,[1] presents the concept of data mining and machine learning which can be used for finding criminal patterns and behaviours. The paper is further divided by providing basic differentiation of the clustering techniques used in unsupervised learning. And then after the crime dataset of India that contains record of serious fraud of property in all states and we will apply k-means clustering to find generic patterns. The main reason for this paper is to give a quip thought of how machine learning can be utilized by the law authorization to distinguish, anticipate and illuminate violations by a lot quicker rate.

M. Saldaña et al.,[2] presents a methodology of analysis of crime facts from online newspapers, identifying the different communes where the greatest number of criminal events occur, which gives an idea of potentially more dangerous places, through the detection and geographical mapping of critical points, or the analysis of the nature of the crime through the extraction of entities. Statistics that measure the predictive capacity of the model indicate that the methodology is robust to recognize crime events within the body of the news.

L. Al-Sahan et al.,[3] provides the Blockchain technologies to facilitate the exchange of relevant surveillance events as admitted transactions into a permissioned Hyperledger fabric Blockchain. We conducted comprehensive analysis to demonstrate the feasibility of blockchain and the efficiency of the machine learning-based face recognition and matching for real-time surveillance of suspects using heterogeneous surveillance infrastructure. The proposed architecture proved scalability and real-time behavior after putting the system through multiple test cases. With very high matching accuracy, and end-to-end latency of less than 12.8 seconds, the system proves to be scalable, and fast enough for a smart surveillance use case.

G. Borowik et al.,[4] show the usefulness of analytic algorithms in predicting crimes, however, there are other applications of such analyzes in the area of law enforcement, such as defining criminal hot spots, creating criminal profiles, and detecting crime trends. The most important factor is the accuracy with which one can infer and create new knowledge based on observations from the past that will be useful in the process of reducing the number of crimes (predictive policing) and ensure the security of citizens.

P. V. Savyan et al.,[5] proposes a method based on unsupervised clustering which analyse the reactions of users called smileys. The reactions are profiled and by applying similarity measures and unsupervised clustering techniques, they are further classified. This approach reveals the behaviour of immediate emotional responses of users to the various posts in Facebook. Since reactions are immediate, the analysis of these reactions provides important information to find anomalous behaviour in Facebook accounts.

S. T. Bharathi et al.,[6] this approach trained the proposed system with supervised data set with collected crime information from various places of Tamil Nadu through online available data. In the testing phase, first identify the cluster closest to the test crime by using K-Medoids clustering algorithm and then identify the suspected criminal list using similarity measure. The initial stage of implementation and analysis of the proposed scheme provides good results and high accuracy. The proposed scheme is compared with related K-Means clustering algorithm with same set of training and test.

J. Kim et al.,[7] presents malicious behavior by tracking the execution flow of binary code. Our method of tracking the execution flow of the binary code utilizing the BFS (Breath-First Search)algorithm advances static analysis based on binary code, but it can be a method combining the advantage of static analysis and the advantage of dynamic analysis. In addition to visualizing malicious behavior as a graph image based on APIs, it is possible to analyze more obviously malicious behavior.

A. Iqbal et al.,[8] shows the studies dealing with the log analysis and correlation of very specialized setups in industrial control systems implemented in the context of power systems. These cases consider the behavior of logs and their ability or inability to shed light on the incriminating nature of a criminal investigation. Our research is novel and unique in the sense that no such previous study exists detailing the forensic investigation on ICS within power sector.

U. Thongsatapornwatana et al.,[9] proposes the criminal behavior analysis method to detect suspect vehicles that are potentially involved in criminal activity. It must not rely on the blacklist. The analysis is conditional on journey path and the involvement of criminal activities. In additional, public officials believe that the suspect vehicle will choose the journey path without a checkpoint. Therefore, we used the journey path analysis techniques together with the association rule mining to analyze such criminal behavior. From extensive experiments, the results show that the proposed method can increase the suspect detection accuracy rate 17.24% beyond the traditional counterpart.

E. E. Hemdan et al.,[10] presents analysis approach for batch and stream log data using Apache Spark. The results show that Spark can be used as a fast platform for handling the diverse large size of log data and extract useful information that can assist digital investigators in the analysis immense amount of generated cloud log data in a given frame of time. Furthermore, the results can make provision to reconstruct and generate a timeline related to historical past sequence events occurred during a cloud crime as well as identify the malicious user's IP address, date and time, with a number of accesses.

J. Liu et al.,[11] The goal of this project is to apply the theory and technique of ontology in analyzing the malware behavior and then to use the result of this study in classifying malware into the forensic-aware categories. Assisted by these proposed categories, prosecutors will have the upper hand to defeat suspect's alibi in a more efficient way. Moreover, the result of this project will help computer forensics professionals to present their expert opinions against malware defense in court with the uniformity and clarity.

S. Smit et al.,[12] present QUIN, a support system that can model different crime scenarios and reason about what happened. For this it uses expert knowledge and historical data. Investigating with QUIN provides a solid overview of current knowledge of the case, a clear view on likely scenarios and suggestions on what to look for in order to advance the investigation.

III. MACHINE LEARNING IN CRIMINAL ACTIVITIES

The machine learning techniques is suitable to predict the criminal activities. Banks, for example, are halting financial crimes much more quickly and cheaply than they used to by using AI for automating processes and conducting multilayered “deep learning” analyses. AI tools have permitted them to shrink the armies of people they employ to evaluate alerts for suspicious activities. That’s because their false alerts have fallen by as much as half thanks to AI, and because many banks are now able to automate routine human legwork in document evaluation. For example, using artificial intelligence, Paypal has also cut its false alerts in half. And Royal Bank of Scotland prevented losses of over $9 million to customers after conducting a year-long pilot with Vocalink Analytics, a payments business, to use AI to scan small business transactions for fake invoices.

Figure 2: Sample of criminal data

AI tools also allow companies to surface suspicious patterns or relationships invisible even to experts. For instance, artificial neural networks can enable employees to predict the next moves of even unidentified criminals who have figured out ways around alert triggers in binary rules-based security systems. These artificial neural networks link millions of data points from seemingly unrelated databases, containing everything from social media posts to internet protocol addresses used on airport Wi-Fi networks to real estate holdings or tax returns, and identify patterns [4].

The next step in assessing the wisdom of launching an AI risk-management program is for companies to evaluate to what extent customers and government authorities will expect them to be ahead of the curve. Even if it does not become a regulatory or legal obligation, companies might find it advantageous to play a leading role in the use of advanced analytics so they can take part in setting industry wide standards. They can help ensure that industry participants, regulators, technology innovators, and customers are being kept safe, without trampling on people’s privacy and human rights.

Increased use of AI tools for crime prevention could also cause external risks to cascade in unexpected ways. A company could lose its credibility with the public, regulators, and other stakeholders in myriad ways -- for example, if there are false alerts that mistakenly identify people as “suspicious” or “criminal” due to a racial bias unintentionally built into the system. Or, at the other end of the spectrum, if they miss criminal activities, like drug trafficking conducted by their clients or funds channeled from

sanctioned countries such as Iran. Criminals could resort to more extreme, and potentially violent, measures to outmaneuver AI. Customers could flee to less closely monitored entities outside of regulated industries. A moral hazard could even develop if employees become too reliant on AI crime-fighting tools to catch criminals for them. Employees could feasibly develop a false sense of comfort, and then stop regularly checking the outputs and miss obvious cases [5].

To prevent this from happening, companies need to create and test a variety of scenarios of cascading events resulting from AI-driven tools used to track criminal activities. To outsmart money launderers, for example, banks should conduct "war games" with ex-prosecutors and investigators to discover how they would beat their system.

With results produced through scenario analysis, managers can then help top executives and board members decide how comfortable they are with using AI crime-fighting. They can also develop crisis management playbooks containing internal and external communication strategies so they can react swiftly when things (inevitably) go wrong.

By using AI, companies can identify areas of potential crimes such as fraud, money laundering, and terrorist financing – in addition to more mundane crimes such as employee theft, cyber fraud, and fake invoices – to help public agencies with prosecuting these offenses much more effectively and efficiently. But with these benefits come risks that should be openly, honestly, and transparently assessed to determine whether using AI in this way is a strategic fit. It will not be easy. But clear communication with regulators and customers will allow companies to rise to the challenge when things go wrong. AI will eventually have a hugely positive impact on reducing crime in the world – as long as it is managed well [6].

IV. CHALLENGES

A. Biological Factors

Early biological theories in criminology took the view that structure determines function- that is, individuals behave differently because of the fundamental fact that they are somehow structurally different. These theories tended to focus strongly on inherited characteristics. Modern biological theories in criminology, in contrast, examine the entire range of biological characteristics, including those that result from genetic defects (and thus are not inherited) and those that are environmentally induced.

B. Family Studies

Explanations of human behavior in terms of heredity go far back in antiquity and are based on the common sense observation that children tend to resemble their parents in appearance, mannerisms, and disposition. In connection with the development of the theory of heredity, new statistical methods were devised to measure degrees of resemblance or correlation Charles Goring used these new statistical techniques in the analysis of criminality, arriving at the conclusion that crime is inherited in much the same way, as are ordinary physical traits and features.

C. Neurotransmitters

Neurotransmitters are chemical that allow for the transmission of electrical impulses within the brain and are the basis for the brain's processing of information. As such, they underlie all types of behavior, including antisocial behavior. About thirty studies have examined the linkage between neurotransmitters and antisocial behavior.

D. Hormones

In addition to neurotransmitter levels, much research has been generated relating to the effect of hormone levels on human behavior, including aggressive or criminal behavior. Interest in hormones dates back to the mid 1800s, when biochemists were first able to isolate and identify some of the physiological and psychological effects of the secretions of the endocrine glands (hormones). Most recent attention paid to hormone levels and aggressive or criminal behavior relates to either testosterone or female premenstrual cycles.

V. PROPOSED STRATEGY

- Load the Criminal Review Dataset from the Kaggle

In this step, the criminal review dataset will be downloaded from kaggle source. It is a large dataset providing company. Then load this dataset into the python environment.

- Visualizing the Dataset

Now open the dataset files and view the various data in term of features like age, sex, time, area, purpose etc.

- Pre-process the Dataset

Now the data preprocess step applied, here data is finalize for processing. Missing data is either removal or replace form constant one or zero in this step.

- Splitting the Dataset into training and testing

In this step, the final preprocessed of dataset is divided into the training and the testing dataset. In the machine learning, firstly the machine is trained through given dataset then it comes in tested period for remaining dataset.

- Classification Using Machine Learning Algorithm

Now apply the machine learning technique to find the performance parameters.

- Performance Metrics

(Accuracy, Precision, Recall, F1 - Score)

Now the performance parameters are calculated in terms of precision, recall, f-1 measure, accuracy etc by using the following formulas-

True Positive (TP): predicted true and event are positive.

True Negative (TN): Predicted true and event are negative.

False Positive (FP): predicted false and event are positive.

False Negative (FN): Predicted false and event are negative.

$$Precision = \frac{|TP|}{|TP| + |FP|}$$

$$Recall = \frac{|TP|}{|TP| + |FN|}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

$$Accuracy = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}$$

VI. CONCLUSION

Crimes are growing rapidly in every states of every country with a much faster frequency rate and police or FBI departments are struggling to obtain some patterns and solving crimes with a vast number of datasets. Therefore, there is a need to identify these crimes based on the pattern recognition and making predictions by applying some data science techniques and methods. The crime rates can be significantly reduced by the real-time crime forecasting and mass surveillance, which are helpful in saving lives that is the most valuable thing. Therefore need to implement and analysis of criminal behavior model based on machine learning.

REFERENCES

[1]. G. Jha, L. Ahuja and A. Rana, "Criminal Behaviour Analysis and Segmentation using K-Means Clustering," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2020, pp. 1356-1360, doi: 10.1109/ICRITO48877.2020.9197791.

[2]. M. Saldaña, C. Escobar, E. Galvez, D. Torres and N. Toro, "Mapping of the Perception of Theft Crimes from Analysis of Newspaper Articles Online," 2020 15th Iberian Conference on Information Systems and Technologies (CISTI), Seville, Spain, 2020, pp. 1-7, doi: 10.23919/CISTI49556.2020.9141154.

[3]. L. Al-Sahan, F. Al-Jabiri, N. Abdelsalam, A. Mohamed, T. Elfouly and M. Abdallah, "Public Security Surveillance System Using Blockchain Technology and Advanced Image Processing Techniques," 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar, 2020, pp. 104-111, doi: 10.1109/ICIoT48696.2020.9089523.

[4]. G. Borowik, Z. M. Wawrzyniak and P. Cichosz, "Time series analysis for crime forecasting," 2018 26th International Conference on Systems Engineering (ICSEng), Sydney, NSW, Australia, 2018, pp. 1-10, doi: 10.1109/ICSENG.2018.8638179.

[5]. P. V. Savyan and S. M. S. Bhanu, "Behaviour Profiling of Reactions in Facebook Posts for Anomaly Detection," 2017 Ninth International Conference on Advanced Computing (ICoAC), Chennai, India, 2017, pp. 220-226, doi: 10.1109/ICoAC.2017.8441402.

[6]. S. T. Bharathi, B. Indrani and M. A. Prabakar, "A supervised learning approach for criminal identification using similarity measures and K-Medoids clustering," 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kerala, India, 2017, pp. 646-653, doi: 10.1109/ICICICT1.2017.8342639.

[7]. J. Kim and J. M. Youn, "Malware behavior analysis using binary code tracking," 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT), Kuta Bali, Indonesia, 2017, pp. 1-4, doi: 10.1109/CAIPT.2017.8320724.

[8]. A. Iqbal, M. Ekstedt and H. Alobaidli, "Exploratory studies into forensic logs for criminal investigation using case studies in industrial control systems in the power sector," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 2017, pp. 3657-3661, doi: 10.1109/BigData.2017.8258360.

[9]. U. Thongsatapornwatana, W. Lilakiatsakun, A. Kawbunjun and T. Boongoen, "Analysis of criminal behaviors for suspect vehicle detection," 2017 Twelfth International Conference on Digital Information Management (ICDIM), Fukuoka, Japan, 2017, pp. 15-20, doi: 10.1109/ICDIM.2017.8244645.

[10]. E. E. Hemdan and D. H. Manjaiah, "Spark-based log data analysis for reconstruction of cybercrime events in cloud environment," 2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT), Kollam, India, 2017, pp. 1-8, doi: 10.1109/ICCPCT.2017.8074209.

[11]. J. Liu, R. Kammar, R. Sasaki and T. Uehara, "Malware Behavior Ontology for Digital Evidence," 2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), Prague, Czech Republic, 2017, pp. 585-586, doi: 10.1109/QRS-C.2017.105.

[12]. S. Smit, B. Van Der Vecht, F. van Wermeskerken and J. W. Streefkerk, "QUIN: Providing Integrated Analysis Support to Crime Investigators," 2016 European Intelligence and Security Informatics Conference (EISIC), Uppsala, Sweden, 2016, pp. 120-123, doi: 10.1109/EISIC.2016.031.